1995

Comparison of three speakers along critical acoustic dimensions

Karin Bramlett

Comparison of Three Speakers Along Critical Acoustic Dimensions

Independent Study

Advisor: Dr. James Miller

by

Karin Bramlett

Spring 1995

Comparison of Three Speakers Along Critical Acoustic Dimensions

Independent Study

Advisor: Dr. James Miller

by

Karin Bramlett

Spring 1995

Speech production is a complicated process involving the neurological system as well as the motions of articulation and breathing. Numerous studies provide information and data about various aspects of normal speech production and the speech of the deaf. However, few of these studies provide information about the speech of elementary-school-aged children. Several features of the speech of two eight-year-old girls, one with normal hearing and the other hearing-impaired, were examined in this study. Although this study involved limited subjects and data samples, the information gained should prove useful to anyone wishing to organize a more in depth study of a similar nature.

Method

Subjects

Three female subjects of differing hearing ability provided data for this study. The first subject was a normal hearing eight-year-old. The second subject, also eight years old, was profoundly deaf having a pure tone average of 116 dB in the right ear and 115 dB in the left. She was born deaf and has worn a cochlear implant for 3 years. The final subject was a hearing-impaired adult whose pure-tone average (500, 1000, and 2000 Hz) was 33 dB. She began losing her hearing at the age of 26 and has worn hearing aids for three years.

Materials

The subjects were asked to read two times each a list of 15 words and three sentences. The words selected included most of the vowels and diphthongs of spoken English. The sentences provided a variety of speech sounds and processes.

<u>Word List</u>

bake, ball, bat, bead, bed, big, bike, bird, boat, body, book, boot, bow-wow, boy, bud.

<u>Sentences</u>

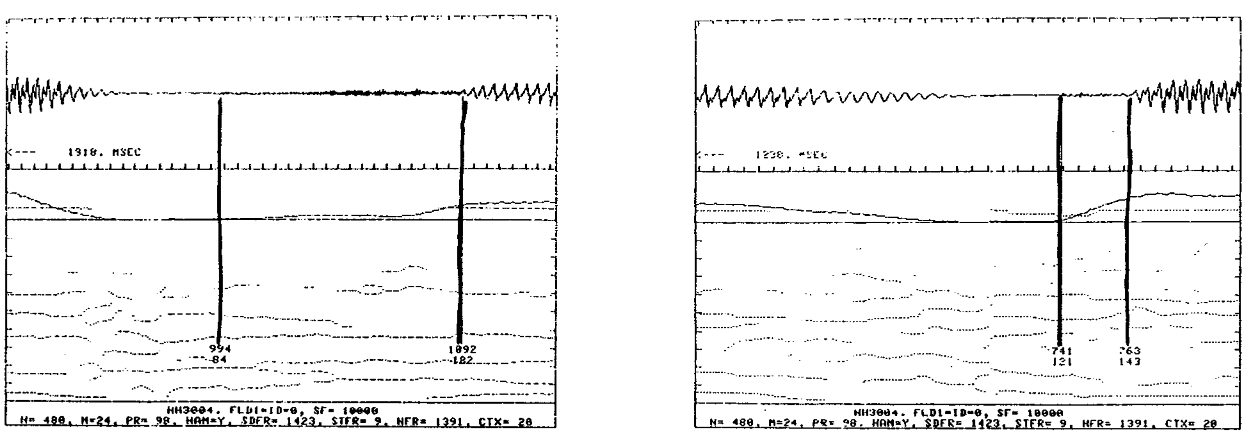
He has a blue pen.
I am tall.
David likes new toys.

The words and sentences were presented to the subjects by the experimenter. A microphone was placed approximately 25 cm from the speaker's lips and speech was recorded on a Sony DAT recorder. Each subject read the list of words and sentences twice. Digitizations of all of the speech samples were made at 20 kHz with 16-bit precision through a Digisound-16, a stimulus access processor, and both a 50 Hz analog high-pass filter (to remove incidental noise) and a 10 kHz anti-aliasing filter. Any residual AC noise or "hum" was removed by digitally notch filtering the files at 60 Hz after digitization. The files were stored on a Micro Vax II or VAXstation 3200 computer to facilitate further processing by the Interactive-Laboratory-System (ILS) commercial software package.

Acoustic Measurements:

For acoustic analysis, the LPC-spectrogram resulting from the SGM command of the ILS package was used.

Voice onset time. Voice onset time (VOT) in plosives refers to the moment at which voicing starts relative to the release of a closure (Ladefoged, 1993). In normal speech, phonation for /b/, /d/, and /g/ begins very shortly after stop release, while there is a delay of 50 ms before phonation begins following the release of /p/, /t/, and /k/ (Borden et al., 1994). The VOT time difference is depicted in Figure 1.



/p/ /b/

Figure 1. Spectrograms of the phonemes /p/ and /b/ produced by the normal hearing subject. The VOT of /p/ = 98 ms; The VOT of /b/ = 22 ms. The difference in length of aspiration = 76 ms.

According to Mackay (1987), computer synthesized syllables produced in small increments falling on a continuum between /pa/ and /ba/ were most often heard as either /pa/ or /ba/ instead of something in between. Because of this, each speaker was compared to herself instead of to other speakers.

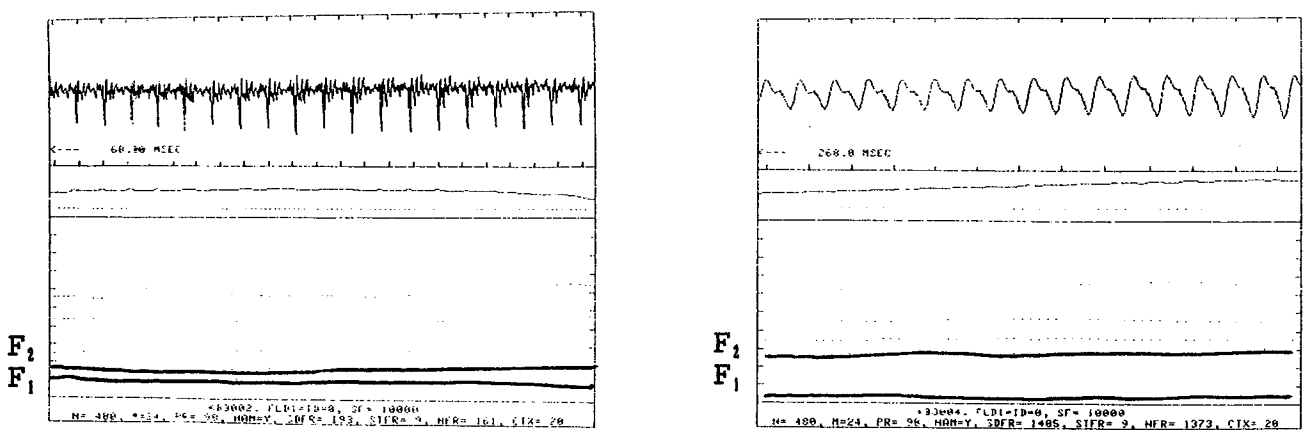
The VOTs of /b/ and /p/ were taken from the sentence "He has a blue pen." The VOT of /t/ was taken from "I am tall," and the VOT of /d/ was taken from "David likes new toys." The duration between the burst and the onset of voicing was measured for each consonant. The VOT of the voiced consonant was subtracted from the VOT of its unvoiced counterpart (i.e. /p/-/b/ and /t/-/d/).

Liquids and Nasals: In his 1978 study, Monsen looked at the production of liquids and nasals. He found two objective characteristics to distinguish normal from deviant productions. In normal speech, he found a clearly marked boundary between the syllable initial nasal or liquid and the following vowel. In much of his data on deaf speakers, he found no clear cut boundary. He also found that the duration of nasals and liquids rarely exceeded 100 ms in normal speech whereas deaf speakers elongated these sounds sometimes greater than 400 ms.

In this study, the liquid /l/ and the nasal /n/ from the sentence "David likes new toys" were analyzed. The experimenter recorded the length of the consonants noting any production over 200 ms as abnormal.

Variables of Vowel Production: Vowels are produced with an open oral tract and consist of different formants. According to the Speech Science Primer (Borden et al., 1994), the first formant (F_1) is most responsive to changes in mouth opening and the second formant (F_2) is most responsive to changes such as tongue placement or tongue movement within the oral cavity. The relationship of these formants to each other determines which vowel the listener perceives.

This study looked at the variability of F_1 and F_2 in vowel production, and the extent of F_2 frequency change in the diphthong /ai/. The variability of F_1 and F_2 were determined by comparing formants from the vowels /a/ and /i/. These were chosen because /a/ has a high F_1 and a low F_2 compared to /i/ which has a low F_1 and a high F_2. These differences are depicted in Figure 2.

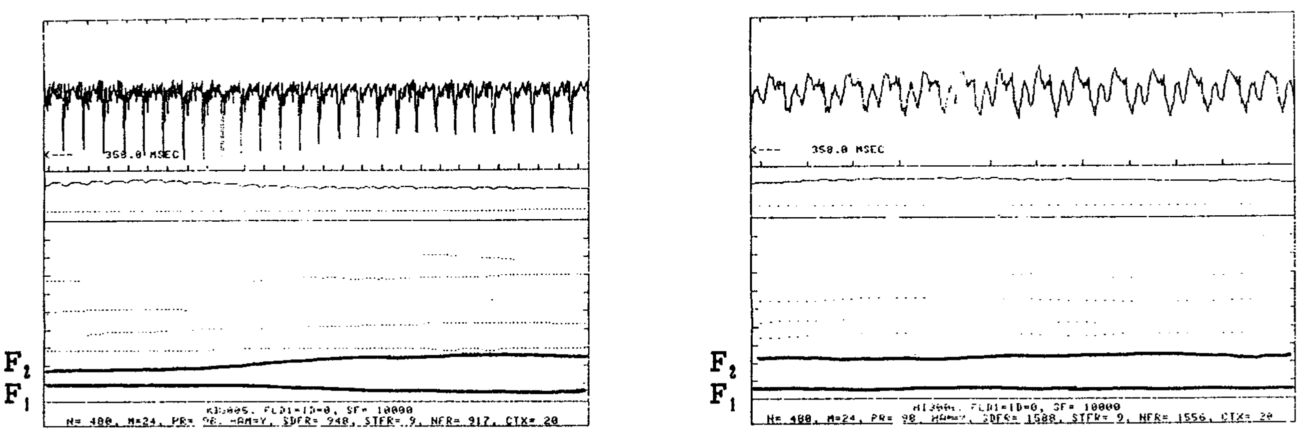


/a/ /i/

Figure 2. Spectrograms of the vowels /a/ and /i/ produced by the adult subject. The formants were sampled in three positions throughout the production and averaged. Averages for /a/: F_1 = 1066 Hz, F_2 = 1839 Hz. Averages for /i/: F_1 = 399 Hz, F_2 = 2836 Hz.

The range of F_1 was determined by subtracting F_1/a/ - F_1/i/ and F_2 was determined by subtracting F_2/i/ - F_2/a/. The data for /a/ was taken from the word "body," and /i/ was taken from the sentence "He has a blue pen."

The diphthong /ai/ is a combination of the vowels /a/ and /i/. The speaker must rapidly close the mouth, spread the lips, and move the tongue forward to produce this sound correctly. This motion lowers F_1 and raises F_2 concurrently. Monsen (1978) found that the most common error in deaf vowel production was limited change in F_2. Figure 3 compares a correct production of /ai/ to a deviant one. Pictured on the left, the correct production shows F_1 decline from 992 Hz to 632 Hz and F_2 ascend from 1818 Hz to 2524 Hz. The deviant production on the right shows little movement in F_1 or F_2. Data for /ai/ was taken from the sentence "I am tall."



Subject: Adult

Subject: Deaf Child

Figure 3. Spectrograms of the diphthong /ai/ in the speech of two subjects; the correct production on the left produced by the hearing impaired adult and the incorrect production on the right produced by the deaf child.

Variables of prosody: Many variables in prosody exist between speakers. In deaf speakers, the duration of words or sentences often seems excessively long to the listener (Monsen, 1978). In this study, the length of each sentence was measured to see what differences, if any, existed between the three subjects.

RESULTS

Voice onset time:

According to Monsen (1978), a difference of 50 ms indicates that the speaker can differentiate between the production of the voiced plosive and its unvoiced counterpart. He found that when hearing-impaired speakers fail to produce a distinction between the voiced and unvoiced plosives /b/ vs. /p/ and /d/ vs. /t/, they tend to produce an unaspirated consonant for both.

As shown by the data in Table 1, the hearing child met the 50 ms standard and the hearing impaired adult fell slightly below in both sets of comparisons.

Voice Onset Time in Voiced and Unvoiced Plosives

Comparison of /t/ and /d/:

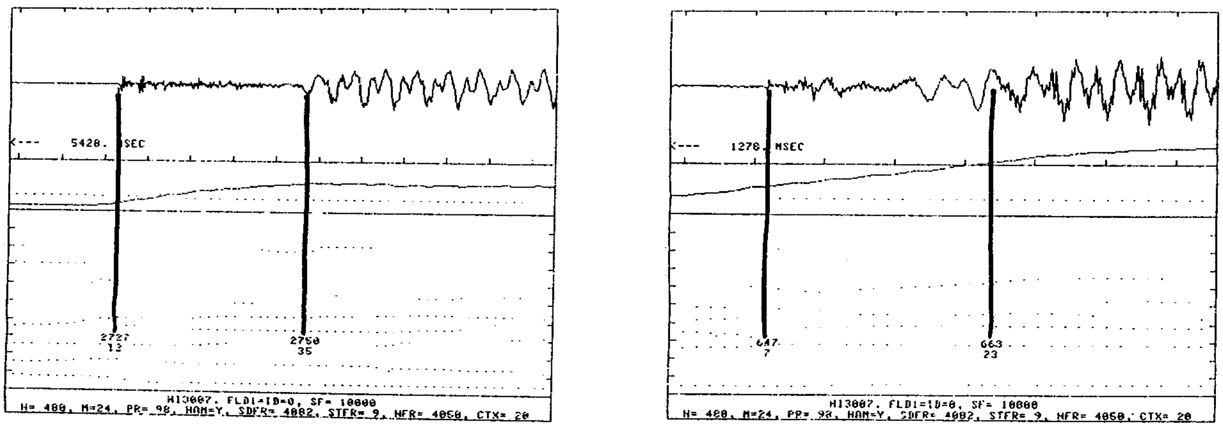
Subject	/t/	/d/	/t/ - /d/	Average
Hearing Child	116	18	98	76 ms
	83	29	54	
Deaf Child	23	16	7	.5 ms
	15	21	-6	
Adult	74	18	56	48 ms
	53	13	40	

Comparison of /p/ and /b/:

Subject	/p/	/b/	/p/ - /b/	Average
Hearing Child	54	23	31	54 ms
	98	22	76	
Deaf Child	14	147	-133	-106 ms
	24	102	-78	
Adult	55	20	35	36 ms
	56	19	37	

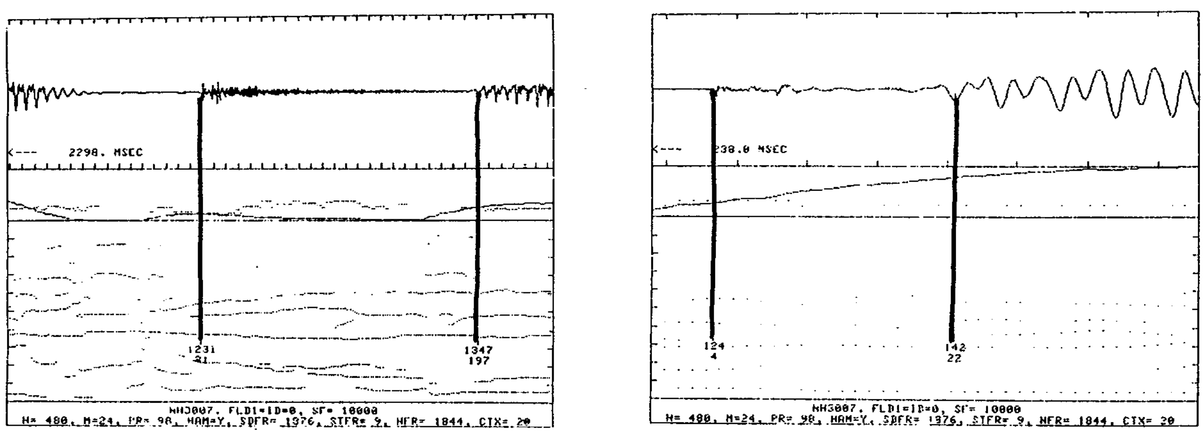
Table 1. Results of VOT measurements.

In the average production of /t/ and /d/, the deaf child made almost no distinction between the duration of the two. However, the spectrograms in Figure 4 show that the phonemes were not produced identically. The /d/ appeared to be an aspirated voiced consonant and the /t/ appeared to be an unaspirated unvoiced consonant. The spectrograms in Figure 5 portray an accurate example of /t/ and /d/ for comparison.



/t/ /d/

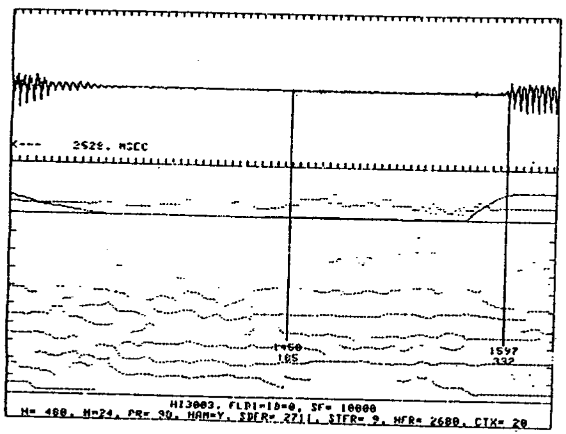
Figure 4. Spectrograms of the phonemes /t/ and /d/ produced by the deaf child. The VOT of /t/ = 23 ms; The VOT of /d/ = 16 ms. The difference in length of aspiration = 7 ms.



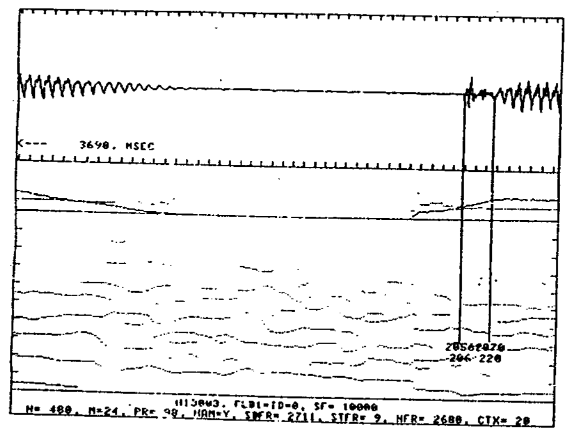
/t/ /d/

Figure 5. Spectrograms of the phonemes /t/ and /d/ produced by the normal hearing child. The VOT of /t/ = 116 ms; The VOT of /d/ = 18 ms. The difference in length of aspiration = 98 ms.

In the production of /p/ and /b/, the deaf child's speech differed in duration by more than 50 ms. However, the subject reversed the phonemes in both readings to produce the sentence, "He has a plue ben" instead of "He has a blue pen." Figure 6 shows the deaf child's normal production of /b/ and an elongated production of /p/. This data can be compared to Figure 1 on page 4 which shows correct productions of /p/ and /b/ by the normal hearing child.



/p/



/b/

Figure 6. Spectrograms of the phonemes /p/ and /b/ produced by the deaf child. The VOT of /p/ = 147 ms; The VOT of /b/ = 14 ms. The difference in length of aspiration = 133 ms.

Nasals and liquids:

The information presented in Table 2 reveals that all productions fell below 200 ms in duration and formed clear boundaries before the following vowels. However, in one of the two samples of deaf speech, the deaf child substituted /j/ for /l/ and deleted the /n/. The substitution might have been caused by an error in constriction. Although the /l/ and /j/ require similar tongue placement, air escapes over the tongue for /j/ but laterally around the sides of the tongue for /l/. This is an understandable error for a deaf speaker to make. Because the contrast in sound is produced by subtle changes within the mouth, the deaf speaker is unable to lipread a difference.

Liquids and Nasals

Duration of /l/ and /n/ in Milliseconds:

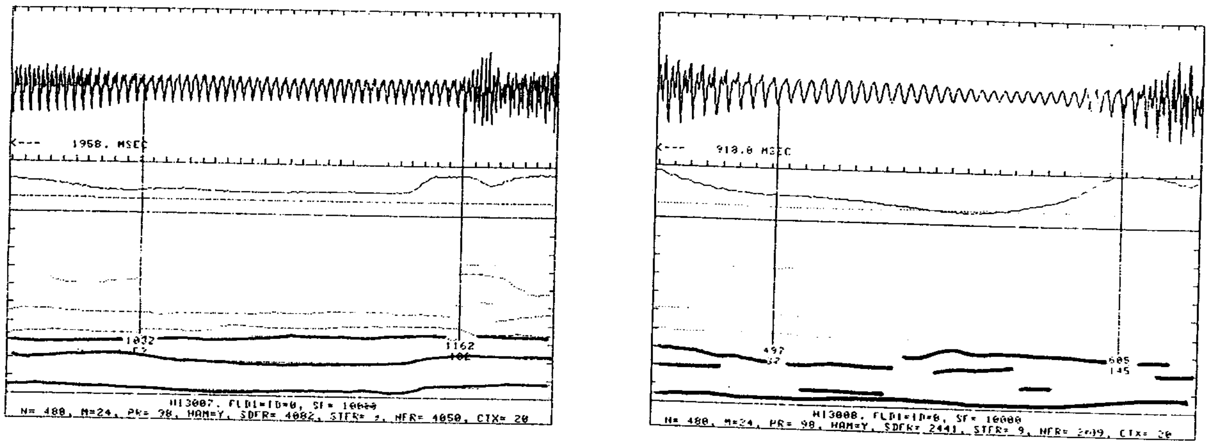
Subject	**/l/**	**/n/**
Hearing Child	96	127
	108	80
Deaf Child	130	65
	0	0
Adult	55	38
	32	47

Clearly Marked Boundary:

Subject	**/l/**	**/n/**
Hearing Child	yes	yes
	yes	yes
Deaf Chid	yes	yes
	substituted /j/	deleted /n/
Adult	yes	yes
	yes	yes

Table 2. Data collected on /l/ and /n/ for all subjects.

Figure 7 shows the deaf child's correct production of /l/ on the left and the substitution of /j/ on the right. The samples were taken from the sentence "Davi<u>d l</u>ikes new toys," and coarticulation can be seen between the /d/ and the /l/ and the /d/ and the /j/. Each illustration demonstrates a clear boundary before the following vowel /ai/.



Correct Production

Substitution of /j/

Figure 7. Spectrograms of the phonemes /l/ and /j/ produced by the deaf child. Duration of /l/ = 130 ms and /j/ = 108 ms.

Variables of Vowel Production

The first and second formants of the vowels /a/ and /i/ were measured and graphed in Figures 8 - 10. Figure 8, generated from the adult subject's speech, shows a large difference between the maximum and minimum values for both formants. This indicates a normal range of vowel production.

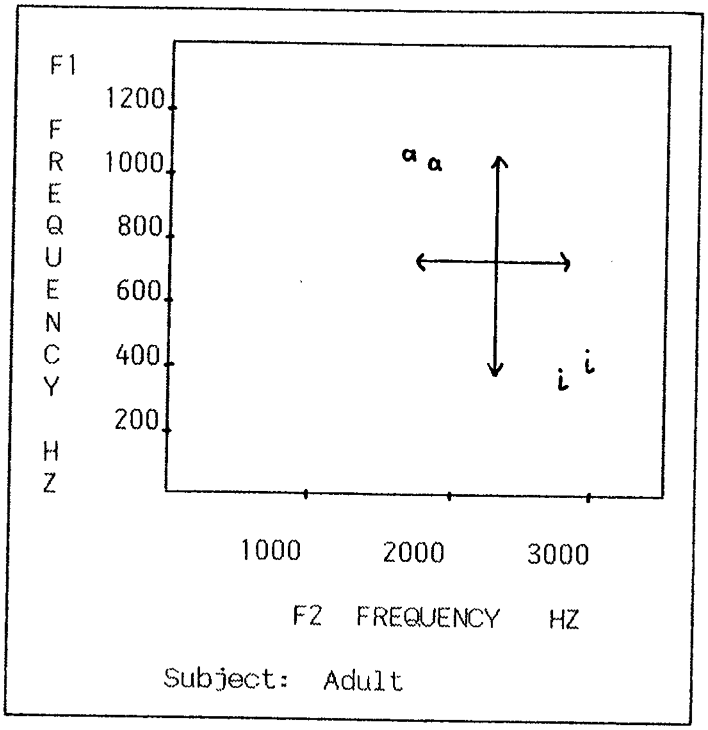


Figure 8. Variability of the first and second formants as indicated by the formant frequencies of the vowels /a/ and /i/. Subject: Adult; approximately normal range of variation of both formants.

The children's speech samples indicated otherwise. The deaf child's speech (Figure 9) shows the vowels collapsed along the dimension of F_2. This data coincides with the findings of Monsen in 1978. Since F_2 is most influenced by changes within the oral cavity, this data indicates that the deaf child had little variation in tongue movement but correctly adjusted the mouth opening for each production.

The normal hearing child's speech showed a different pattern. Figure 10 shows the vowels collapsed along the dimension of F_1. This indicates that the child used proper tongue placement but did not widen her mouth opening when producing /a/.

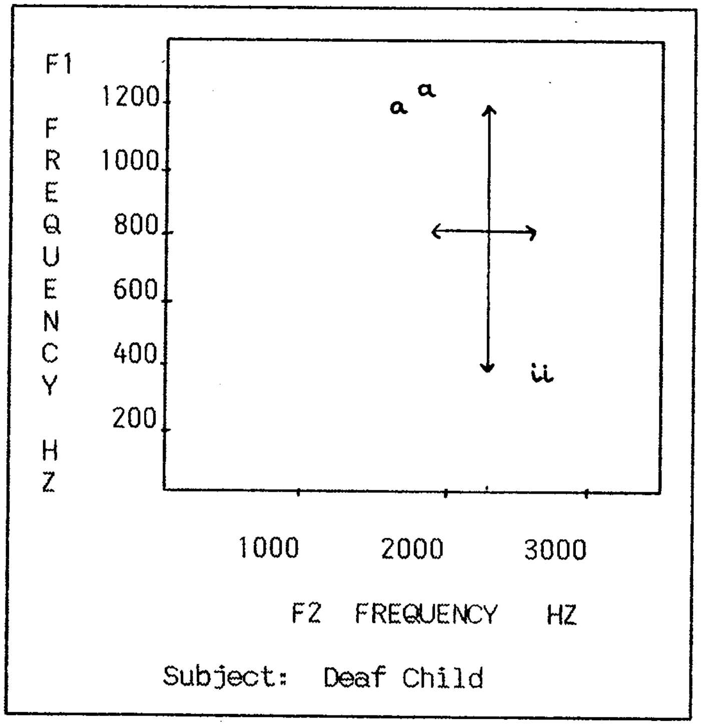


Figure 9. Variability of the first and second formants as indicated by the formant frequencies of the vowels /a/ and /i/. Subject: Deaf Child; restricted range of second formant.

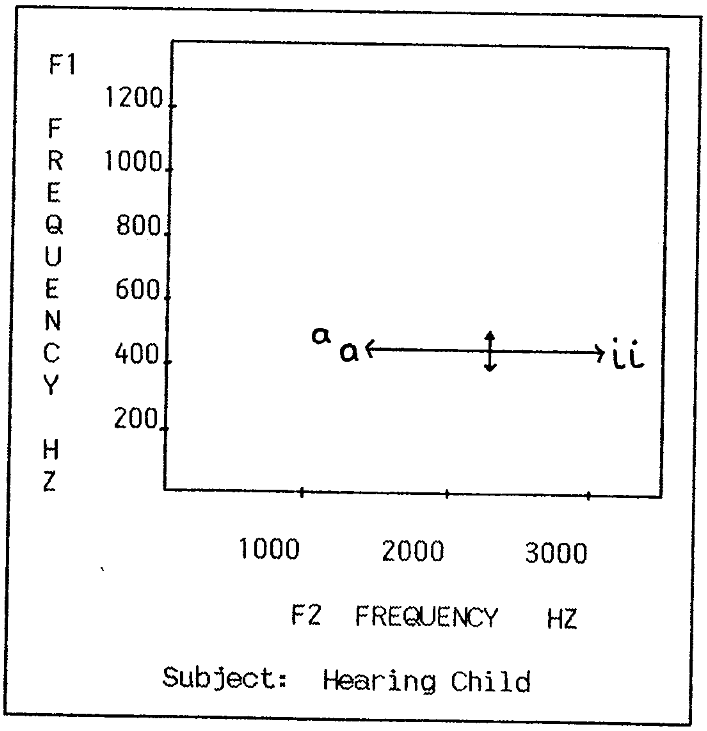


Figure 10. Variability of the first and second formants as indicated by the formant frequencies of the vowels /a/ and /i/. Subject: Hearing Child; restricted range of first formant.

The range of variation was calculated and recorded in Table 10. Figure 11 depicts the average range of variation of the first and second formants of the vowels /a/ and /i/ for each subject.

Variability of First and Second Formants

Range of First Formant:

Subject	F1 /a/	F1 /i/	/a/ -/i/	Average
Hearing Child	506	467	39	14 Hz
	453	464	-11	
Deaf Child	1251	391	860	833 Hz
	1207	401	806	
Adult	1088	466	622	645 Hz
	1066	399	667	

Range of Second Formant:

Subject	F2 /i/	F2 /a/	/i/ - /a/	Average
Hearing Child	3358	1171	2187	2054 Hz
	3282	1361	1921	
Deaf Child	2572	1947	625	734 Hz
	2633	1791	842	
Adult	2944	1609	1335	1166 Hz
	2836	1839	997	

Table 3. Data used to calculate the range of variation in F_1 and F_2.

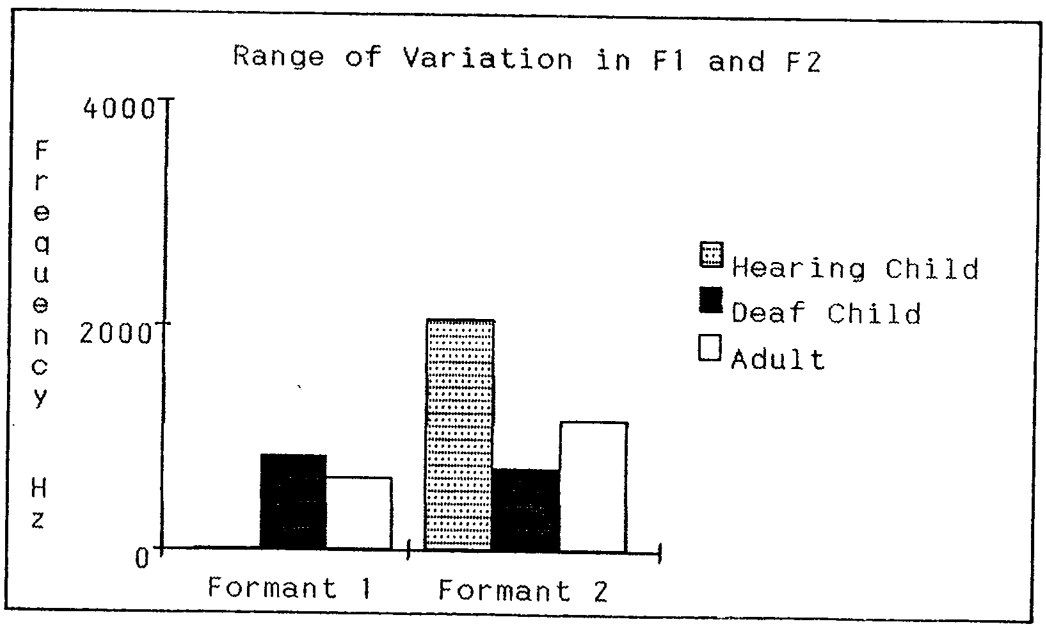
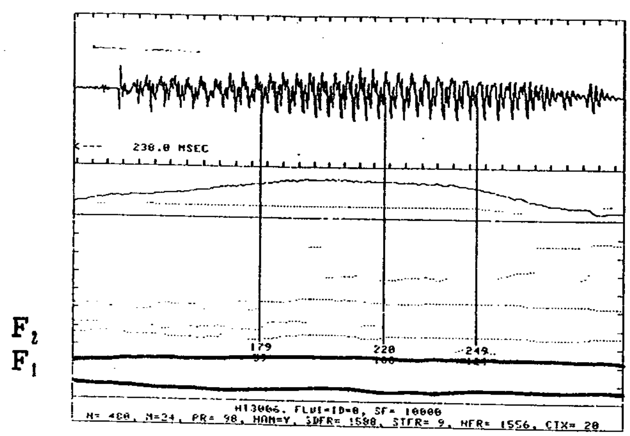


Figure 11. The average range of variation of F_1 and F_2 for each subject.

The deaf child failed to produce adequate variation in F_2 across the vowels /a/ and /i/. A more difficult task was producing the diphthong /ai/ where the formants must change rapidly. A spectrogram of the deaf child's /ai/ is shown in Figure 12. The first and second formant remain constant indicating that the speaker did not change the tongue or mouth positions during the diphthong. Although the formants remained constant, the first formant of /ai/ (634 Hz) fell between the deaf child's first formants /i/ (396 Hz) and /a/ (1229 Hz). Likewise, the second formant of /ai/ (2342 Hz) fell between the second formant of /a/ (1869 Hz) and /i/ (2603 Hz).

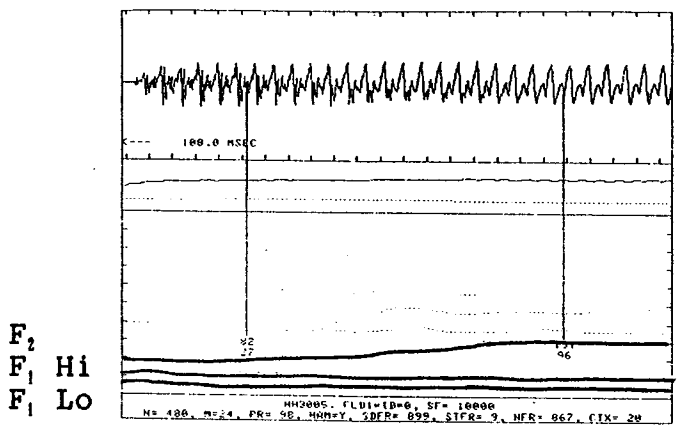


/ai/

Figure 12. Spectrogram of the diphthong /ai/ produced by the deaf child. Average F_1 = 634 Hz; Average F_2 = 2342 Hz. Both formants show little variation.

In Figure 3 on page 7 the spectrogram of the adult's speech shows normal variation in both formants; F_1 falls 360 Hz as F_2 increases 706 Hz.

The hearing child showed restricted variation in F_1 between the vowels /a/ and /i/. This pattern remained constant in the production of /ai/; F_2 increased 1136 Hz while F_1 showed little movement. As seen in Figure 13, an additional first formant was created in this production of /ai/. According to Miller (1989), this additional formant could be caused by nasality. Since /ai/ was taken from the sentence "I am tall," the speaker may have nasalized the /ai/ in anticipation of the nasal /m/.



/ai/

Figure 13. Spectrogram of the diphthong /ai/ produced by the hearing child. F_1 shows limited variation while F_2 shows normal variation. F_1 Low and F_1 High may indicate nasality.

Variables of Prosody

The length of each sentence per speaker is listed in Table 4 and the averages per speaker are charted in Figure 14. Monsen (1978) states that deaf speech often seems longer than normal speech. In this study, the duration of the deaf child's speech was longer than the hearing child's in two out of three of the sentences. Sentence 1 was 66% longer and Sentence 3 was 59% longer. Sentence 2 was approximately the same length for both speakers.

Duration of Sentences		
"He has a blue pen."		
Subject	**Duration**	**Average**
Hearing Child	1449	1409 ms
	1369	
Deaf Child	2559	2333 ms
	2107	
Adult	1378	1321 ms
	1263	
"I am tall."		
Subject	**Duration**	**Average**
Hearing Child	750	753 ms
	756	
Deaf Child	785	761 ms
	737	
Adult	745	767 ms
	789	
"David likes new toys."		
Subject	**Duration**	**Average**
Hearing Child	1693	1655 ms
	1617	
Deaf Child	2916	2625 ms
	2333	
Adult	1606	1535 ms
	1463	

Table 4. Sentence length in milliseconds of three sentences.

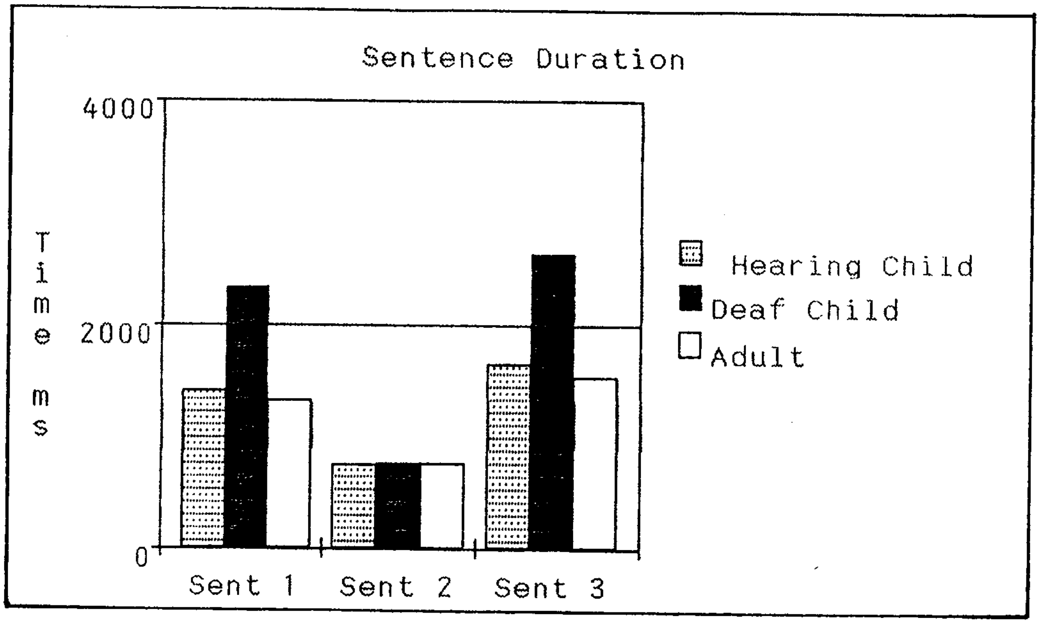


Figure 14. Average duration of sentence per speaker. Sentence 1: "He has a blue pen." Sentence 2: "I am tall." Sentence 3: "David likes new toys."

DISCUSSION

Since only one hearing-impaired adult, one deaf child, and one hearing child were utilized in this study, no theories can be proven or refuted here. However, the results can be compared to the findings of Monsen in his 1978 study and new questions can be raised.

Monsen suggested that a 50 ms time difference demonstrated a speakers ability to distinguish between voiced and unvoiced plosives. The data from this study showed that the normal hearing child made that distinction whereas the hearing-impaired adult nearly made the distinction and the deaf child failed to make the distinction in the case of /t/ and /d/. This confirms the well-known fact that the accuracy of plosive production is related to the speaker's hearing level.

Monsen also targeted the production of liquids and nasals. Although the deaf child in this study produced the /l/ and /n/ correctly in one reading of the sentence, she substituted /j/ for /l/ and deleted /n/ in the next. This indicates that even though the deaf speaker was able to produce liquids and nasals properly, she could not do so reliably.

Formant information from both the adult subject and the deaf child agreed closely with the data taken by Monsen. The data collected on the hearing child, however, was most unexpected. Further study in this area could determine how the hearing child's vowel production compared to that of other normal hearing children. It might be found that a large change in F_1 is not necessary for the production of an adequate /a/-/i/ distinction.

In this study, information was obtained that confirms Monsen's statement that deaf speech often sounds elongated. The speech of the deaf child was significantly longer than that of the hearing child in two out of three of the test sentences.

Further research in the field of speech production is needed by speech therapists and teachers of the deaf. By determining errors of production and their causes, researchers can provide teachers with greater insight on the complicated process of speech production. This insight in turn may help teachers better educate their students.

REFERENCES

Borden, G.J., Harris, K.S., and Raphael, L.J., *Speech Science Primer; Physiology, Acoustics, and Perception of Speech.* Baltimore: Williams and Wilkins, (1994).

Ladefoged, P., *A Course in Phonetics.* United States: Harcourt Brace Jovanovich, Inc., 142 (1993).

Mackay, I., *Phonetics: The Science of Speech Production.* Boston: Little, Brown, 291 (1987)

Miller, J.D., Auditory-perceptual interpretation of the vowel. *Journal of the Acoustical Society of America*, 85, 2127 (1989).
b

Monsen, R.B., Toward measuring how well hearing-impaired children speak. *Journal of Speech and Hearing Research*, 21, 197-219 (1978).